

The impact of fine size and uncertainty on punishment and deterrence: Theory and evidence from the laboratory

Eberhard Feess

Frankfurt School of Finance & Management

Hannah Schildberg-Hörisch*

DICE, University of Düsseldorf and IZA

Markus Schramm

RWTH Aachen University

Ansgar Wohlschlegel

Portsmouth Business School

Abstract

Increasing punishment is typically considered first choice to boost deterrence of unwarranted behavior such as false financial statements, asset misappropriation, stealing, or corruption. However, if there is uncertainty on a potential violator's guilt, judges' and juries' willingness to impose punishment may decrease in its magnitude. Thus, increasing the magnitude of punishment may backfire, when the reduced punishment probability is anticipated by potential violators. Based on a theoretical model, our paper is the first to analyze the interdependency of violation and punishment behavior in a laboratory experiment, and to contrast it to the standard partial equilibrium perspective on deterrence that considers the punishment probability to be independent of the fine size. Varying both the magnitude of fines and the degree of uncertainty shows that, in case of legal uncertainty, the deterrent effect of higher fines is far less pronounced than if the punishment probability was exogenous.

JEL Classification: K14; K42; C91

Keywords: deterrence; punishment; uncertainty; Blackstone Ratio; partial and equilibrium effects; laboratory experiment

*Corresponding author: Düsseldorf Institute for Competition Economics (DICE), Heinrich-Heine-University Düsseldorf, oeconomicum 24.31, Universitätsstr. 1, 40225 Düsseldorf, Germany, email: schildberg-hoerisch@dice.hhu.de, phone: +49 221 8114318. Declaration of interest for all authors: none. Financial support from the German Research Foundation (DFG) through SFB-TR 15 and by the Portsmouth Business School's Research Project Funds is gratefully acknowledged and did not influence study design or analysis of results. We thank Anne Mertens for programming and excellent assistance in implementing the experiment and participants of EEA 2016 in Geneva, the Youth Crime and Public Policy Interventions Conference in Surrey, and the Workshop on Economics and Finance at Karlsruhe Institute of Technology for helpful comments.

1 Introduction

Illegal behavior imposes an enormous burden on society. Examples for such behavior range from asset misappropriation by employees and false financial statements by companies to corruption in bureaucracies or any other kind of criminal behavior. At the company level, estimated losses from asset misappropriation by own employees add up to about 5 percent of businesses' annual revenues (Association of Certified Fraud Examiners, 2014). The overall costs of crime and justice including defense measures such as alarm systems are tremendous (Cohen, 2005). As a consequence, institutions that reduce the occurrence of illegal behavior are key to the wealth and well-being of nations (Acemoglu and Johnson, 2005).

In his seminal paper on crime and punishment, Gary Becker (1968) pointed out that deterrence of criminal behavior is increasing in the expected punishment, and many empirical studies suggest that crime rates are in fact decreasing in fines and in the probability of apprehension (e.g. Levitt, 1998; Haselhuhn et al., 2012; DeAngelo and Hansen, 2014; Hansen, 2015; see Chalfin and McCrary, 2017, for a recent review of the literature). The deterrent effect of higher fines, though, is only straightforward when fines and conviction probabilities can be seen as independent choice variables. Otherwise, high fines may backfire when they violate the fairness considerations of judges, who may then be reluctant to convict a defendant.

The willingness to convict suspects, however, does not only depend on the perceived adequacy of fines, but also on the degree of uncertainty that someone is actually guilty. In case of uncertainty, decision makers know that they can make two mistakes, convicting innocent defendants (usually referred to as type I errors) and releasing guilty defendants (type II errors). As, for a given level of uncertainty, any reduction in type I errors leads to a certain incline of type II errors, the conviction probability in case of actual guilt depends on the relative weight judges or juries put on these two error types. This weight in turn depends on the interplay of fine size and uncertainty as many people might be unwilling to impose fines in general and large fines in particular when they have strong doubts about a defendant's guilt. If this effect is strong enough and anticipated by potential violators, higher fines may even lead to lower deterrence (Andreoni, 1991; Feess and Wohlschlegel, 2009). It is the effect of this interdependency of fines and uncertainty on the frequency of punishment and illegal behavior that we are interested in.

To the best of our knowledge, we are the first to develop a theoretical model and a laboratory experiment to distinguish between the *partial* and *total effects* of fines and uncertainty: For partial effects, we follow the overwhelming part of the literature which considers the impact of fines and uncertainty under the simplifying assumption that the violation frequency (when considering judges) or the punishment frequency (when considering violators) is given. Our theoretical analysis confirms the intuition that higher fines reduce both the violation and the punishment frequencies, while higher uncertainty increases the violation frequency, but reduces the punishment frequency.

Results are less straightforward, however, when considering total effects. To see this, suppose that judges are aware that higher uncertainty increases the violation incentives. This, in turn, increases the judges' willingness to punish compared to the case in which the number of violators is given. Hence, total effects differ from partial effects not only with respect to their size, but possibly even with respect to their sign. Our theoretical model allows to fully capture and compare partial and total effects. For total effects, we find that a higher fine size leads to a lower punishment frequency, whereas the total impact of the fine size on the violation frequency is ambiguous. Higher legal uncertainty leads to a higher violation frequency, but the impact of uncertainty on the punishment frequency is ambiguous. Thus, for the impact of the fine size on the violation frequency and for the impact of uncertainty on the punishment frequency, it cannot be taken for granted that partial and total effects go in the same direction.

Based on our theoretical model, we design a laboratory experiment that allows distinguishing between the partial and total effects of legal uncertainty and the magnitude of fines. Due to a lack of suitable field data, we are not aware of any empirical analysis of the interdependency of violation and punishment decisions, so that conducting a laboratory experiment seems particularly useful in this context. In our experiment, participants are randomly assigned to one of two roles, V ("violators") and J ("judges"), and pairs consisting of one participant V and one participant J are matched. For each pair, there is a fixed amount of money designated for donation to a charity. First, participant V decides whether or not to take this money for himself. If it is not taken, it may still disappear due to a random event, and this creates uncertainty. Then, each participant J observes whether the money is still available or has disappeared. Finally, each participant J decides whether or not to punish his randomly assigned participant V if and only if he observes

that the money has disappeared. Note that with uncertainty participant J cannot be sure whether the money has been “stolen” by participant V or disappeared randomly.

Each participant makes four different decisions characterized by combinations of a fine (high or low) and a degree of uncertainty (high or zero). To distinguish between partial and total effects, we employ a strategy method design in which players of one type can condition their behavior on the behavior of the other type. In a first treatment, players J condition their punishment behavior on the violation frequency, while players V decide unconditionally whether to take the money. In the second treatment, players V condition their violation behavior on the punishment frequency, and players J decide unconditionally whether to punish. We then calculate partial as well as total effects: For instance, the partial effect of the fine size on player J ’s punishment behavior is calculated by comparing the punishment frequencies that would have occurred if the violation frequencies were the same for different fine sizes. For the total effect, we compare the punishment frequencies for the different fine sizes and the corresponding *actual* violation frequencies.

Since the frequencies of type I and type II errors are determined endogenously by the decisions of players V and J , the interdependency of violation and punishment decisions is captured by our experiment. In particular, for a given probability that the money disappears without being taken, the risk of making a type I error in case of punishment is strictly decreasing in the number of actual violators. Analogously, the risk of making a type II error in case of no punishment increases in the violation frequency. Thus, for a given punishment frequency, each number of violators translates into a probability for error types.

Our experimental results are fully in line with the theoretical predictions. For partial effects, we confirm that higher fines reduce both the violation and the punishment frequencies. Uncertainty increases the number of violations and decreases the punishment frequency. As outlined above, total effects are theoretically less straightforward, but our empirical findings are again in line with our model: First, a higher fine leads to a significant reduction in the punishment frequency if and only if there is legal uncertainty. Second, the total deterrent effect of the fine size is less pronounced than the partial effect, since players J punish less often when fines are higher. Third, legal uncertainty leads to a pronounced increase in the violation frequency even when violators correctly anticipate the impact on the judges’ punishment frequency. This reinforces the view that legal uncertainty is

worrisome not only from a justice point of view, but also from a deterrence perspective. Fourth, the total impact of legal uncertainty on the punishment frequency is lower than the partial effect, and this holds for both fine levels. Intuitively, introducing legal uncertainty increases the violation incentives, which reduces the incentive to punish less in response to uncertainty. Last and perhaps most importantly, comparing partial and total effects shows that differences can be sizable, especially for changes in the magnitude of fines. Our data on participants' beliefs about the behavior of the other player type show that participants anticipate changes in the other type's behavior in reaction to changes in fines or uncertainty rather well. Thus, taken together, our results underline the need for economic models on deterrence and legislation that take total effects into account.

Concerning the related theoretical literature, we build on and extend the work of Andreoni (1991) and Feess and Wohlschlegel (2009). Both papers argue that potential violators anticipate lower conviction probabilities when fines are higher and, thus, may be more inclined to break the law.¹ Our model extends this literature by allowing judges to take the potential violators' choices rationally into account, and by analyzing comparative statics of the violation and the conviction frequencies with respect to fine size and level of uncertainty.

We go beyond existing empirical evidence in several respects. In order to test Becker's (1968) deterrence hypothesis in the laboratory, Harbaugh, Mocan, and Visser (2013), Khadjavi (2015), Rizzolli and Stanca (2012), and Schildberg-Hörisch and Strassmair (2012) use designs with exogenous punishment schemes (fine sizes and probabilities of punishment) in which participants can "steal" from their counterparts' endowments. Further studies test the deterrent effect of exogenous punishment schemes using designs that mimic corruption (Abbink, Irlenbusch and Renner, 2002), cheating (Nagin and Pogarsky, 2003), (non-)compliance with regulation (Friesen, 2012), or roadway speeding (DeAngelo and Charness, 2012). Except for DeAngelo and Charness (2012) who frame participants' decisions as a choice whether or not to speed, all papers use a context-free, neutral framing. Overall, these papers confirm that the violation frequency is decreasing in fines and the probability of punishment.² In contrast, our paper analyzes endogenously arising punishment

¹Other countervailing equilibrium effects of higher punishment have been identified by Huck and Kosfeld (2007) who point out that citizens' willingness to participate in crime prevention may be crowded out if the deterrent effect of punishment is too strong and Franzoni (1999) who argues that higher punishment reduces prosecutors' incentives to investigate.

²Moreover, studying both context-free and framed treatments, Anderson et al. (2017) study a two-period enforce-

patterns by taking the interdependency of violation and punishment decisions into account.

So far, most experimental papers on endogenous punishment have adopted voluntary contribution mechanisms (VCMs), in which participants can mutually sanction non-cooperative behavior, building on the work of Fehr and Gächter (2000). Legal fines resemble a situation with third-party punishment, in which those who can punish are not directly affected by the behavior of violators. Fehr and Fischbacher (2004), Anderson and Putterman (2006), Carpenter (2007), and Leibbrandt and López-Pérez (2011) study third-party punishment in settings like dictator games or VCMs but neither consider uncertainty nor the interdependence of violation and punishment. Baumann and Friehe (forthcoming) also analyze total effects of fines and legal uncertainty in an experiment but, as they do not use a strategy-method design, they cannot compare total to partial effects. They also find that potential violators are more likely to “steal” when legal uncertainty is high. However, neither the punishment nor the violation probability reacts to different levels of sanctions.

According to the famous Blackstone ratio it is “better that ten guilty persons escape, than that one innocent suffers”. Most people share the view that avoiding type I errors matters more, but the marginal rate of substitution between both error types varies considerably (Volkh, 1997). Using a take game, Rizzolli and Stanca (2012) show experimentally that the adverse effects of type I errors on deterrence are higher than those of type II errors. In their setting, the probabilities of the two errors types are given exogenously, i.e. they do not consider the behavior of participants in the role of judges. Our experimental results show that participants, on average, care far more about type I errors, and that the willingness to take the risk of type I errors decreases in the fine size.

Finally, while uncertainty increases the violation frequency in our model and experiment, Harel and Segal (1999) point out that uncertainty on the punishment probability enhances deterrence if it entails ambiguity. The experiment by DeAngelo and Charness (2012) lends support to the benefits of ambiguity on deterrence, while Salmon and Shniderman (2017) find almost no impact. In contrast to these papers, subjects in our setting may also be punished without taking the money, which reduces the benefit from obeying the rule. In addition, uncertainty over causation reduces the judges’ willingness to punish, which reinforces the detrimental impact of uncertainty on deterrence.

ment regime and show that fine structures that are decreasing over time are typically more effective at deterring repeated risky behaviour than those that are increasing over time.

2 A model of punishment and deterrence

We develop a model that allows analyzing the impact of fine size and uncertainty on the interplay between punishment and deterrence. Basic models on errors in court (Png, 1986, Polinsky and Shavell, 2000) find that both type I and type II errors increase the violation frequency as the difference between the probability of being punished with and without violation shrinks. A full-fledged equilibrium analysis, however, needs to take into account that higher uncertainty and higher fines may reduce judges' willingness to punish, so that the relative probabilities for the two error types need to be derived endogenously.

We consider a two-sided game of incomplete information with two players, a violator and a judge. Violating the law yields a monetary payoff A for the violator which is identical to the victim's loss. In case of theft, for instance, the infringement effectively creates a redistribution from the victim to the defendant.³ Without infringement, an exogenous event may still cause the same loss to the victim with probability q (for instance, she may misplace her wallet). Our assumptions imply that the loss can occur only once, which follows models on uncertainty over causation in the tradition of Shavell (1985). If the loss has occurred for whichever cause, the potential violator is tried in court. Without observing whether the defendant had infringed the law,⁴ the judge then decides whether or not to impose an exogenously given fine F on the defendant.⁵ If the defendant is punished and had infringed the law, the gain from crime is restituted, i.e. in addition to fine F , A is paid back to the victim.⁶

Infringing the law imposes a disutility equal to m_j on violator j , which differs among violators. Potential violators observe their own types m_j , but judges only observe the distribution $H(m)$ over the whole population. Suppose a potential violator believes to be punished with probability p when

³Our assumption that gains and losses from crimes are identical is realistic in some but not in all cases. Adding an efficiency loss due to the infringement would not change any of the model's qualitative results, but it would add another motive to punish for efficiency seeking judges. This would make it more difficult to empirically isolate the impact of uncertainty and fine size and their interaction in our experiment.

⁴As in all models on errors in court, we need to assume that the facts of the case cannot be fully reconstructed ex post.

⁵Our model assumes that there is a two-tier system in which decision makers can only decide about conviction based on a predetermined fine size. While this is often realistic, there are also cases in which juries can convict and advocate low sentences.

⁶Again, the extent of restitution varies widely across cases in reality, but does not influence any of the model's qualitative results.

infringing the law. Then, his expected monetary payoff is $(1 - p)A - pF$ as he can keep A if and only if he is not punished. As he faces type-specific preference costs m_j when violating the law, his expected net utility is

$$U_j^V = (1 - p)A - pF - m_j,$$

where superscript “V” denotes “violation”. Without infringement, his expected utility is

$$U_j^N = -qpF,$$

where superscript “N” denotes “no violation”. It follows that a violator j will infringe the law if and only if his preference costs are below a specific threshold \tilde{m} , i.e. if and only if

$$m_j \leq \tilde{m} := A(1 - p) - Fp(1 - q). \quad (1)$$

For judges, we introduce the following assumptions: First, we set the utility from correct decisions to zero, that is, we take only the preference costs of misjudgments, but not the benefits of correct decisions into account. Second, we assume that judges are heterogenous in the way they take the consequences of their decisions for others into account. Specifically, the disutility from a type I error for judge i is assumed to be $\alpha_i F$ where α_i is a parameter on judge i ’s aversion against type I errors, and F is the fine size. Likewise, a judge’s disutility from a type II error is $\beta_i A$, where A is the violator’s benefit from infringement (and the harm imposed on others by the adverse outcome), which we will use as a measure for the severity of the infringement. As all that matters for the judge’s choice between conviction and acquittal is the ratio of the degrees of aversion against type I and type II errors, we normalize $\beta_i = 1$, so that α_i captures this ratio. A judge i observes her preference parameter α_i . However, the potential violator does not know which type α_i of judge he will be matched with if he is tried but knows its ex-ante probability distribution $G(\alpha)$.

Recall that, when a defendant is tried, a judge cannot observe whether the adverse outcome was due to an infringement by the defendant or due to an exogenous event. However, in a Perfect Bayesian Equilibrium, she will form beliefs that are consistent with equilibrium strategies. In particular, if the unconditional probability that an arbitrary potential violator infringes the law is ϕ (the ‘violation frequency’), then judge i ’s ex-post belief that a violation took place after observing the adverse outcome is $\frac{\phi}{\phi + (1 - \phi)q}$, and the ex post probability of no infringement is $\frac{(1 - \phi)q}{\phi + (1 - \phi)q}$. Thus,

$\pi = \frac{(1-\phi)q}{\phi+(1-\phi)q}$ is the judge's probability of committing a type I error when punishing, whereas $1 - \pi = \frac{\phi}{\phi+(1-\phi)q}$ is the probability of a type II error in case of no punishment. Hence, a judge i who believes that the observed adverse outcome is due to the defendant's infringement with probability ϕ prefers to punish the defendant if and only if

$$\frac{(1-\phi)q}{\phi+(1-\phi)q} \alpha_i F \leq \frac{\phi}{\phi+(1-\phi)q} A, \quad (2)$$

which is equivalent to

$$\alpha_i \leq \tilde{\alpha} = \frac{A\phi}{Fq(1-\phi)} = \frac{1-\pi}{\pi} \frac{A}{F}. \quad (3)$$

The first equation in (3) shows that punishment is more likely if judges expect a higher violation frequency ϕ . The reason is that the probability of a type I error when punishing, π (a type II error when not punishing, $1 - \pi$) is strictly decreasing (increasing) in the violation frequency. In other words, the punishment frequency as an increasing function of the violation frequency can be transformed into a decreasing function of π , the probability of making a type I error, which is displayed in the second equation in (3). This demonstrates that, when taking changes of the violation frequency into account, the well-known discussion of the impact of type I and type II errors on judges' punishment behavior turns into a discussion of the impact of violation frequencies on punishment.

Equations (3) and (1) characterize the judges' and violators' optimal decisions based on their beliefs. Inspecting these two equations yields the following Proposition:

Proposition 1 *Suppose that violators treat the punishment frequency p , and judges the violation frequency ϕ , as exogenously given. Then:*

- (i) $\frac{\partial \tilde{m}}{\partial p} = -A - F(1-q) < 0$, $\frac{\partial \tilde{m}}{\partial F} = -p(1-q) < 0$, and $\frac{\partial \tilde{m}}{\partial q} = Fp > 0$ and
- (ii) $\frac{\partial \tilde{\alpha}}{\partial \phi} = \frac{A}{Fq(1-\phi)^2} > 0$, $\frac{\partial \tilde{\alpha}}{\partial F} = -\frac{A\phi}{F^2q(1-\phi)} < 0$, and $\frac{\partial \tilde{\alpha}}{\partial q} = -\frac{A\phi}{Fq^2(1-\phi)} < 0$.

Part (i) of Proposition confirms the well-known result that violators are less likely to violate if they assume a higher punishment probability p and the fine size F is larger. Furthermore, the violation frequency increases in the probability q that the loss occurs also without infringement.⁷

⁷Our results are in line with the widespread view that increases in the conviction probability are a more effective

Part (ii) discusses judges' behavioral responses to the model parameters: When they assume violators to break the law more frequently, when the evidence on the actual punishment is less noisy, and when fines are low, then judges will also punish for higher aversion against type I errors, so that $\tilde{\alpha}$ increases. In a way, $\frac{\partial \tilde{m}}{\partial p}$ characterizes how violators' best responses depend on the judges' punishment frequency as, for any given p , a potential violator steals if and only if $m \leq \tilde{m}$. Similarly, $\frac{\partial \tilde{\alpha}}{\partial \phi}$ captures the effect of the assumed violating frequency on judges' best responses.

While these individual best responses are intuitive, things are more complicated when we take the interdependency of the decisions into account. In Bayesian Nash equilibrium, judges' and violators' beliefs about their counterparts' actions are consistent with their actual equilibrium choices, i.e. punishment and violation frequencies are given by the system of equations

$$p = G(\tilde{\alpha}) \tag{4}$$

$$\phi = H(\tilde{m}), \tag{5}$$

where $G(\cdot)$ and $H(\cdot)$ denote the cumulative distribution functions of judges' parameter α of relative aversion against type I errors and potential violators' aversion against wrongdoing, provided that the thresholds $\tilde{\alpha}$ and \tilde{m} are in the supports S_α of $G(\cdot)$ and S_m of $H(\cdot)$. The following proposition summarizes the total effect of legal uncertainty and the fine size on equilibrium punishment and violation frequencies:

Proposition 2 *Suppose that S_α and S_m are intervals. If $\tilde{\alpha} \in S_\alpha$ and $\tilde{m} \in S_m$, then the Bayesian Nash equilibrium has the following comparative static properties:*

(i) *For all distributions $G(\cdot)$ and $H(\cdot)$, $\frac{dp^*}{dF} < 0$ and $\frac{d\phi^*}{dq} > 0$.*

(ii) *The absolute equilibrium effects of legal uncertainty on the punishment frequency, and of the fine size on the violation frequency, are smaller than the absolute partial effects derived in Proposition 1.*

deterrent than higher sentences. Usually, this is explained by the psychological fact that, for any expected fine given, people care more about probabilities than about fines (see e.g. Durlauf and Nagin, 2011). In our model, this follows from the larger impact on the expected loss as a percentage reduction of \tilde{m} due to a 1% increase in the expected penalty pF is always greater when this increase is achieved by increasing the conviction probability. The reason is that the former also increases the probability that the donation must be returned.

Proof. See Appendix A. ■

Part (i) of the Proposition shows that, for all distributions of the judges' and the potential violators' preferences, two of the results derived for given behavior of the other party carry over to the Bayesian Nash Equilibrium: First, the punishment frequency decreases in the magnitude of fines, $\frac{dp^*}{dF} < 0$. The intuition is that the consequences of a type I error are always increasing in the fine size, which reduces the judges' willingness to punish. Second, the standard result that higher uncertainty reduces deterrence, $\frac{d\phi^*}{dq} > 0$ also holds in the Bayesian Nash Equilibrium.

However, the total effects of the fine size on the violations frequency ($\frac{d\phi^*}{dF}$) and of legal uncertainty on the punishment behavior ($\frac{dp^*}{dq}$) are ambiguous, and hence do not appear in the Proposition. The first ambiguity confirms the results of Andreoni (1991) and Feess and Wohlschlegel (2009) that higher fines may reduce deterrence if judges care sufficiently strongly about type I errors.⁸ The second ambiguity means that judges may even increase their punishment frequency in response to higher legal uncertainty when the violation frequency increases sufficiently steeply in legal uncertainty.

Part (ii) of the Proposition says that, whenever the total effects are ambiguous, they are quantitatively lower than the partial effects. This is intuitive, since the countervailing effects that make the total effects ambiguous are not present when the behavior of the other party is exogenous. Summing up, our model illustrates how judges' aversion against type I and type II errors causes the punishment frequency in equilibrium to fall in the fine size when there is legal uncertainty.

3 Experimental design

Our objective is to analyze how the judges' punishment and potential violators' decisions react to fines, legal uncertainty, and the behavior of the other type. In order to test the predictions of our model in a comprehensive manner, our design has to ensure that we can compare behavior under the assumption that participants of one type take the other type's decisions as given (partial effects, see Proposition 1) to a situation in which participants anticipate the other type's reaction

⁸Note that in Andreoni (1991), the potentially negative impact of higher fines on deterrence is driven by assumptions on nonlinear preferences, and in Feess and Wohlschlegel (2009) by assumptions on probability distributions of informative signals on the defendant's guilt. By contrast, we assume that judges rationally take into account violators' equilibrium choices, such that this effect arises even for linear preferences and a very simple formalization of legal uncertainty.

to changes in fines or legal uncertainty (total effects, see Proposition 2). For that reason, we employ a strategy method design in which participants of one type can condition their behavior on the behavior of the other type. In two treatments, we vary between subjects whether judges or violators use the strategy method, while the other type decides unconditionally whether to punish or to violate, respectively. Moreover, we vary the level of fine and legal uncertainty within subject: each participant faces four decision situations that are characterized by different combinations of high or low levels of fines and legal uncertainty.

3.1 Punishment and violation decisions

First, participants were randomly assigned the role of potential violators (participants V) or the role of judges (participants J) with equal probability. The instructions referred to potential violators as participants of type A and judges as participants of type B , a fine as “deduction points” and a violation as “taking” instead of stealing. The Online Appendix to this paper contains a translation of the instructions of one treatment to English. Next, pairs consisting of one participant V and one participant J were matched randomly. Throughout the experiment, assigned roles and pairs remained constant and participants acted anonymously. Each participant V and J had an initial endowment of 2400 points. At the end of the experiment, points were converted into Euro at the exchange rate 500 points = 1 Euro. Moreover, for each pair an additional 2400 points were designated for donation to the charity “SOS-Kinderdorf” (SOS Children’s Villages), a very well-known NGO in Germany. The work of the charity was described as follows: “SOS-Kinderdorf supports children and families in the poorest regions of the world (medical care, school enrollment etc.), in crisis areas, but also neglected children and adolescents in Germany. SOS-Kinderdorf has been granted the DZI-certificate that attests reliability and low administrative costs.” In order to model a violation decision, each participant V had to decide whether to take the donation designated for the charity for himself. Taking the donation has real negative consequences for children in need, which makes this act clearly immoral and, thus, for players J creates a motive for punishment that is similar to that of real-world judges in charge of punishing norm violations. Kirchler, Huber, Stefan, and Sutter (2016) and Casal, Fallucchi, and Quercia (2016) use the appropriation of money designated for donation to a charity in a similar way.

Participant J observed whether the donation was available or had disappeared. Importantly, whenever J observed that the donation had disappeared, this could be due to two possible reasons: first, participant V had taken the donation. Second, participant V had decided not to take the donation but it had disappeared by chance, which happened with probability q . The parameter q varies across decision situations and captures legal uncertainty. Only if J observed that the donation had disappeared, he could punish V with a predetermined fine F . If V had taken the donation but was not punished, there was no donation to the charity and his payoff increased by the “donation amount”, 2400 points. If V had taken the donation and was punished, the donation was donated to the charity. If V was punished although he had not taken the donation, the donation remained unavailable and was not donated. If fine F was imposed on participant V , the F points were subtracted from V ’s payoff and additionally donated to the charity. Donating the fine to the charity is important as it avoids potential concerns of participant J that punishing reduces overall efficiency (defined as the sum of payments to participant V , J , and the charity). Figure 1 was part of the participants’ instructions and summarizes all possible decisions, events, resulting situations, and final payments.

Irrespective of the outcome, participant J received 2400 points, but when the donation had disappeared, his decision affected participant V ’s payoff and the donation to charity. We consciously decided not to make participant J ’s payment dependent on the frequencies of type I and type II errors for several reasons. Most importantly, the question whether individuals have different degrees of aversion towards type I and type II errors can hardly be analyzed when exogenously and arbitrarily assigning the same (or different) incentives for avoiding these errors. Moreover, judges in reality cannot be incentivized according to type I and type II errors, simply because there would be no such errors if they were observable. Our questionnaire data suggest that, even without monetary incentives, participants J still suffer strongly from making errors. Using a Likert scale from 0 (not bad at all) to 9 (extremely bad), we asked participants how bad they personally feel if someone innocent is punished for something he has not done (type I error) and if someone guilty is not punished (type II error). The mean answers are 8.2 and 7.5, respectively.⁹

⁹In the experiment by Feess and Sarel (2018), many subjects in the role of judges invest part of their money in order to reduce the probability of type I and type II errors even without own monetary benefits. Moreover, refraining from providing monetary incentives for certain player types has proven useful in related contexts such as Coffman (2011) for a third-party punisher and Schram and Charness (2015) for an advisor on social norms.

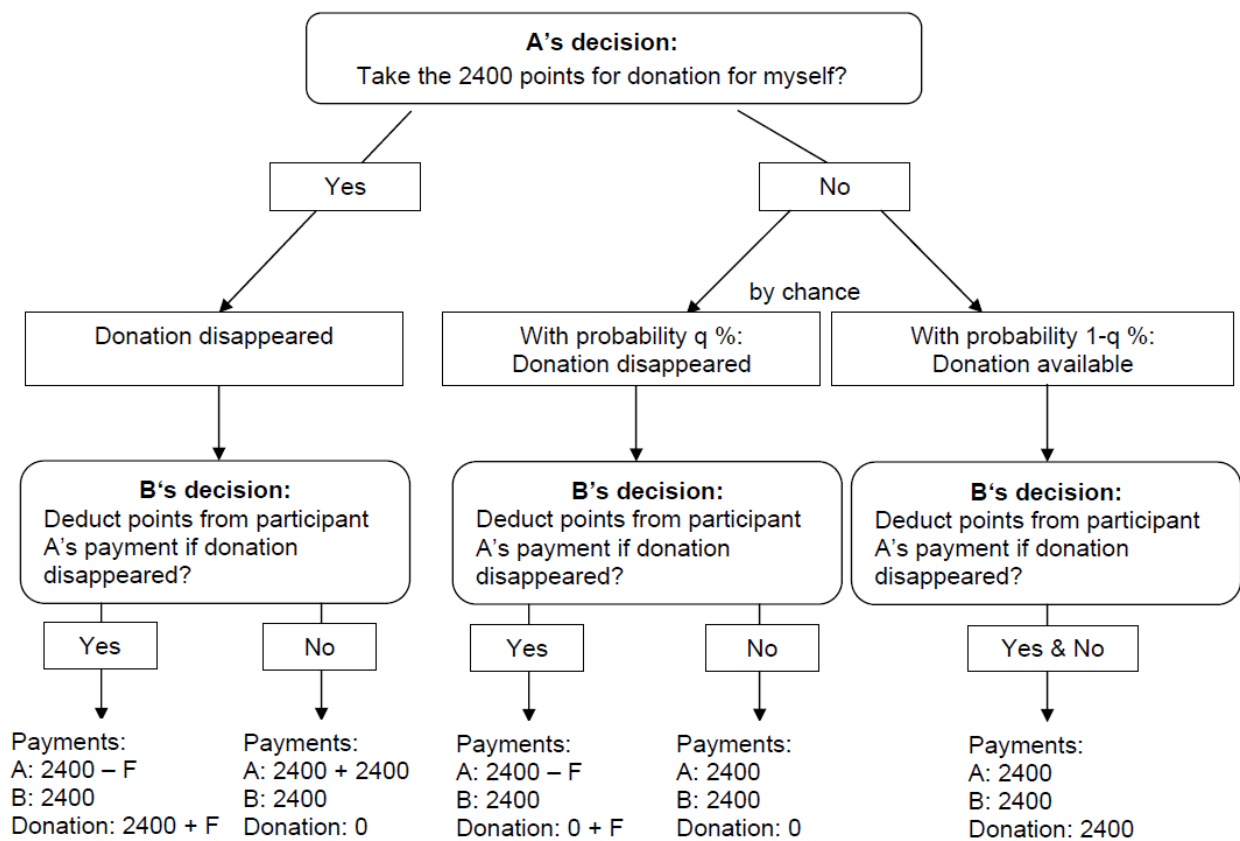


Figure 1

Each participant V and J faced four decision situations that differed only in the size of the fine F ($F = 1200$ or $F = 3600$) and / or the extent of legal uncertainty q ($q = 0$ or $q = 0.5$) in the following order: $F = 1200$ and $q = 50\%$; $F = 3600$ and $q = 50\%$; $F = 1200$ and $q = 0$; $F = 3600$ and $q = 0$ (see Table 1). We set one of the fine sizes above the potential gain from taking the money, as fines in reality often exceed the gain from the violation. Comparing behavior across decision situations enables us to isolate the effect of changes in the fine for a given level of legal uncertainty and vice versa.¹⁰ The instructions described the decision situation in general terms. The exact size of the fine and the level of uncertainty were only revealed on screen just before making the respective decision.¹¹ One of the four decision situations was chosen randomly and the actual decisions of participants V and J in this decision situation determined their earnings and whether there was a donation to the charity.¹²

Table 1: Treatment overview

Treatment	Decision situation			
	1	2	3	4
Player J strategy method (3 sessions, N=72, 36 J , 36 V)	$F = 1200, q = 0.5$	$F = 3600, q = 0.5$	$F = 1200, q = 0$	$F = 3600, q = 0$
Player V strategy method (3 sessions, N=72, 36 J , 36 V)	$F = 1200, q = 0.5$	$F = 3600, q = 0.5$	$F = 1200, q = 0$	$F = 3600, q = 0$

To compare partial and total effects, we used a strategy method design. As shown in Table 1 in half of the six sessions participants J used the strategy method for punishment decisions. For each of the four combinations of fine and legal uncertainty, they made 13 conditional, binary punishment decisions, namely whether their matched participant V should be fined if the donation

¹⁰We consider one situation with legal uncertainty and one without because preferences about error making do not play a role when there is no uncertainty. Hence, the benchmark situation with $q = 0$ informs us about punishment decisions when judges' aversion to error making is not relevant. In this sense, our main aim when varying the level of legal uncertainty is not to analyze comparative statics w.r.t. q but rather a comparison of situations with and without uncertainty. When comparing behavior with and without legal uncertainty we have to assume that the use of the strategy method impacts the decisions of players J in the same way in the simpler environment without legal uncertainty and in the environment in which they face legal uncertainty and thus have to reduce compound lotteries.

¹¹However, the instructions used the decision situation with $F = 1200$ and $q = 0.5$ as an example to illustrate participants' decision screens and carefully explained the implications of legal uncertainty, using the case with $q = 0.5$ as an example.

¹²If the high fine of $F = 3600$ points was imposed on participant V he earned -1200 points that were subtracted from his show-up fee of 2000 points.

had disappeared and x of the 12 participants V in the room had decided to take the donation, for each $x \in \{0, 12\}$. At the same time, participants V made one unconditional violation decision for each of the four decision situations, i.e. they decided whether to take the donation or not. Counting the number of violators in the current session determined which of the conditional punishment decisions was actually implemented.

Similarly, in the remaining three sessions participants V made their violation decisions using the strategy method. In each decision situation, they made 13 conditional, binary violation decisions, namely whether to take the donation if y of the 12 participants J in the room decided to punish if the donation had disappeared, for each $y \in \{0, 12\}$. At the same time, participants J made an unconditional punishment decision for each of the four combinations of fine and legal uncertainty, i.e. they decided whether to punish their randomly matched player V if they observed that the donation had disappeared. The number of punishing participants J in the current session determined which of player V 's conditional violation decisions was implemented.

3.2 Measurement of beliefs

Using the strategy method data, we analyze total effects of changes in fine or legal uncertainty under the assumption that participants correctly anticipate the other type's reaction to the change under consideration. To assess the validity of this assumption, we additionally measure strategy method participants' beliefs on the behavior of the other type. For each of the four decision situations, participants J were asked to guess the number of participants V who had decided to take the donation among the 12 participants V present in the room. Similarly, participants V were asked for their belief on the number of participants J among the 12 present participants J who decided to impose the fine if they observed that the donation had disappeared. Participants knew that their belief for one of the four decision situations would be drawn randomly and paid according to the belief's precision. Participants had the possibility to win additional 2000 points or 4 Euro. The probability of winning increased in the precision of their belief and equaled $p_{win} = 1 - 1/144$ (belief - actual number)². For those participants who were interested in the exact payment scheme, the scheme was explained on screen. This procedure of eliciting beliefs, the randomized quadratic scoring rule, follows Schlag and van der Weele (2013) and incentivizes participants to report their

beliefs truthfully, without imposing assumptions on their risk preferences.

3.3 Procedural details and implementation

The detailed sequence of events in each session was as follows:

1. Instructions: Upon entering the lab, subjects drew a card containing a number and were asked to sit in the respective booth. They read the instructions, were encouraged to ask questions in private, and answered several control questions on the computer.
2. Four decision situations concerning punishment or violation behavior
3. Measurement of beliefs concerning the behavior of the other player type
4. Measurement of risk attitudes¹³
5. Questionnaire¹⁴

The experiment was conducted at the BonnEconLab, Germany with 144 participants, 72 in each treatment. Most participants (94%) were students from various subjects. Age varied between 18 and 57 years (median 23, 92% in the range 19–30 years). We used the software z-Tree (Fischbacher, 2007) for conducting the experiment and ORSEE (Greiner, 2015) for inviting participants from the BonnEconLab’s participant pool and recording their participation. The experiment took place in 6 sessions in June 2015. The experiment lasted about 60 minutes. On average, participants earned 14.20 Euro. Payments consisted of the sum of up to four payments: a show-up fee of 4 Euro, the payment from one randomly drawn decision situation, the incentivized belief (if applicable), and the payment from one randomly drawn risk-related decision. Payments were made in a separate room to ensure privacy. We donated a total of 676.80 Euro to the NGO “SOS-Kinderdorf”, which corresponds to the amount that participants had decided to donate.

¹³We apply a procedure similar to Holt and Laury (2002). Participants had to decide whether they preferred a safe payment of 1200 points or a lottery that would either pay 3000 with probability p or 0 points with probability $(1 - p)$. In ten separate decisions, p was 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1, respectively. One of the ten decisions was drawn randomly and paid.

¹⁴The post-experimental questionnaire measured socio-economic and demographic characteristics, it assessed participants’ general attitudes concerning moral behavior, donations to charity, individual responsibility for own actions, aversion against type I and type II error, and judgments concerning determinants of the effectiveness of legal rules.

4 Hypotheses and results

In our model, we have derived two kinds of results. The *partial effects* summarized in Proposition 1 capture the impacts of fine size F and legal uncertainty q on the punishment frequency for a *given* violation frequency, and the impact on the violation frequency for a *given* punishment frequency. Section 4.1 presents hypotheses and results on partial effects. By contrast, Proposition 2 refers to the *total effects* of changes in fine size or uncertainty, in which reactions to changes in fine size or uncertainty of the other side are correctly anticipated. Section 4.2 compares results on partial and total effects. We additionally report data on beliefs that allow assessing how closely reported beliefs mirror actual behavior.¹⁵

4.1 Partial effects of fine size and uncertainty

4.1.1 Punishment frequency

Following Proposition 1, we test three hypotheses on the punishment frequency of players J :

H1. The punishment frequency increases in the violation frequency.

H2. For identical violation frequencies and with legal uncertainty, the punishment frequency is lower for $F = 3600$ than for $F = 1200$.

H3. For identical violation frequencies and fines, the punishment frequency is lower for $q = 50\%$ than for $q = 0$.

To test these Hypotheses, we use the data for players J in which they condition their punishment behavior on the number of violating players V . In line with H1, the punishment frequency is weakly increasing in the number of violators for 94% or 135 out of 144 decision profiles (36 players J made conditional punishment decisions in 4 decision situations). Figure 2 summarizes player J s' conditional punishment frequency for each of the four decision situations. Each line corresponds to one decision situation and displays the share of players J who punishes (on the vertical axis) for each possible number of violators (horizontal axis). For instance, the solid black line for the decision situation with high fine ($F = 3600$) and high uncertainty ($q = 50\%$) shows that about 21%

¹⁵In an Online Appendix to this paper, we use the data on unconditional punishment and violation decisions to calculate total effects, and compare the results to those derived with the strategy method. The results of this robustness check confirm that the total effects for unconditional decisions are qualitatively the same as those obtained with the strategy method.

of players J punish if two players V take the money, but about 33% of players J punish in case three players V take the money. All players J who punish for two violations also do so for three.¹⁶

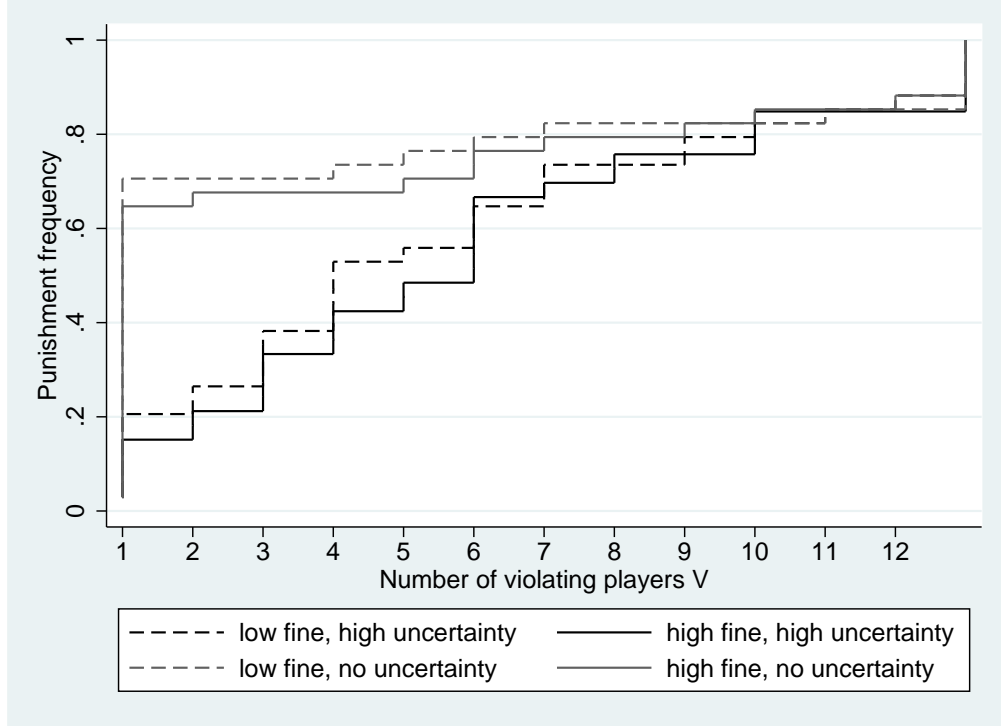


Figure 2: Punishment frequency as a function of the number of violating players for the four decision situations.

Significant differences ($p < 0.05$, signed ranks Wilcoxon tests) between low fine, high uncertainty versus high fine, high uncertainty; low fine, high uncertainty versus low fine, no uncertainty; high fine, high uncertainty versus high fine, no uncertainty; the difference between low fine, no uncertainty and high fine, no uncertainty is not significant ($p = 0.35$, signed ranks Wilcoxon test).

To test H2, the partial effect of an increase in the fine size on judges' behavior is calculated under the assumption that the violation frequency is the same for both fine sizes. Thus, we compare the punishment frequencies for the low and the high fine in case there is high legal uncertainty, $q = 50\%$ (dashed black line versus solid black line). As hypothesized in H2, a significantly higher share of players J punishes if the fine is low instead of high (Wilcoxon signed ranks test, $p = 0.03$; throughout the paper, we report two-sided p -values).¹⁷

¹⁶Figure 2 illustrates the 135 out of 144 decision profiles, for which the punishment frequency is weakly increasing in the number of violators. The tests of Hypotheses 2 and 3 refer to these consistent decision profiles. Test results are qualitatively the same when we consider all 144 decision profiles.

¹⁷Some players J do not punish even in the absence of legal uncertainty (gray lines in Figure 2), which cannot

Moreover, in line with H3, Figure 2 illustrates that for a given fine size, punishment frequencies are significantly lower if legal uncertainty is high: the dashed black line is well below the dashed gray line and the solid black line is well below the solid gray line. Holding constant the level of the fine, a Wilcoxon signed ranks test that compare the punishment frequencies for no and high legal uncertainty yields $p < 0.001$ for both $F = 1200$ and $F = 3600$, respectively.

Moreover, our design allows studying whether individuals have different degrees of aversion towards type I and type II errors, and whether type I error aversion is indeed increasing in the fine size as postulated by our model. Note that, for q given, any exogenously given number of violators translates into error types: for instance, for $q = 50\%$ and five violators (i.e. a violation frequency of $\phi = 5/12$), the probability of a type I error in case of punishment is $\pi = \frac{(1-\phi)q}{\phi+(1-\phi)q} = \frac{7}{17} = 41.2\%$, while the probability of a type II error in case of no punishment is $1 - \pi = 58.8\%$. Substituting $q = 50\%$ and taking into account that the violation frequency is $\phi = V/12$, one can easily check that the frequencies of both error types are identical for four violators; i.e. $\pi = 1 - \pi = 0.5$ for $V = 4$. That implies that players J who already start punishing if there are less than four violators are willing to take a higher risk of making a type I error than making a type II error. Figure 2 illustrates that this is true for 38.2% of players J for $F = 1200$ and for 33.3% for $F = 3600$. By contrast, for both fine sizes, about 50% players J only start punishing if there are more than four violators, so that their decisions imply a larger probability of a type II error. Moreover, the type I error aversion is moderately increasing in the fine size. For $F = 1200$, 53% players J have $V \leq 4$ and 47% players J have $V > 4$, while 42% of judges have $V \leq 4$ and 58% $V > 4$ for $F = 3600$.

4.1.2 Violation frequency

We test the following three hypotheses concerning the violation frequency of players V that directly follow from Proposition 1 and refer to partial effects:

H4. The violation frequency decreases in the punishment frequency.

H5. For identical punishment frequencies and with legal uncertainty, the violation frequency is lower for $F = 3600$ than for $F = 1200$.

H6. For identical punishment frequencies and fines, the violation frequency is higher for $q =$

be explained by preferences against error making as in our theoretical model. Instead, this behavior may reflect a decision maker's reluctance to impose fines perceived as inappropriately high.

50% than for $q = 0$.

In the following, we use the data for players V in which players V condition their violation decisions on the number of punishing players J . As predicted by H4, the violation frequency is weakly decreasing in the number of punishing players J for the vast majority of our subjects, for 91% or 131 out of 144 decision profiles. Figure 3 shows the share of actual violators (vertical axis) for each possible number of punishing players J (horizontal axis).¹⁸ Each of the four lines depicts one of the four decision situations.

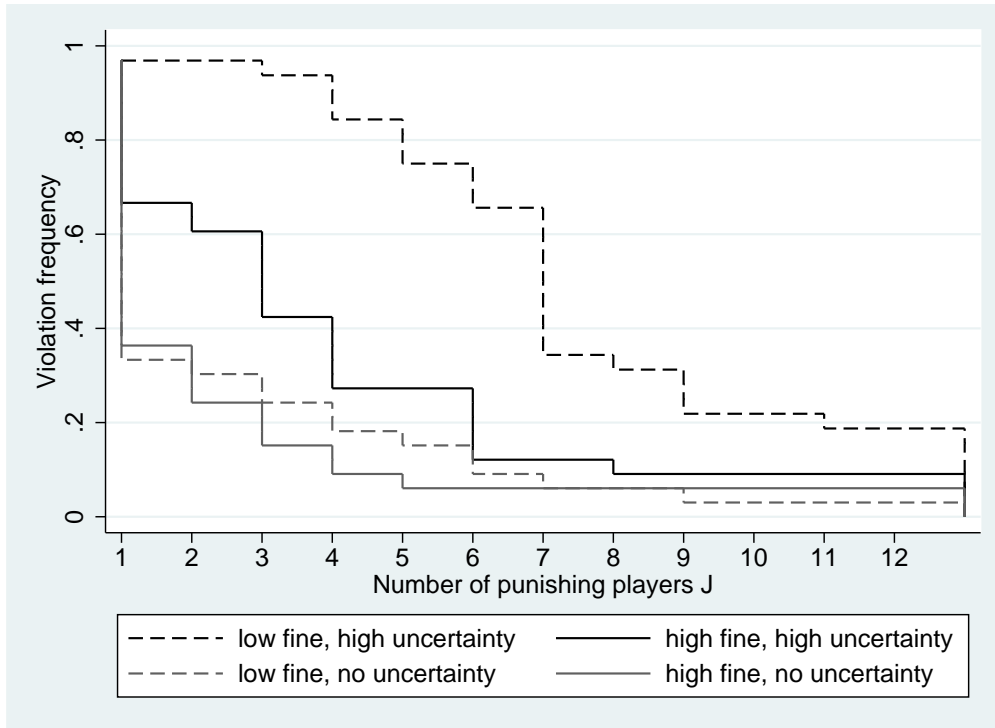


Figure 3: Violation frequency as a function of the number of punishing players for the four decision situations.

Significant differences (two-sided $p < 0.001$ in signed ranks Wilcoxon tests): low fine, high uncertainty versus high fine, high uncertainty; low fine, high uncertainty versus low fine, no uncertainty; high fine, high uncertainty versus high fine, no uncertainty; no significant difference (two-sided $p - values = 0.191$, signed ranks Wilcoxon test): low fine, no uncertainty versus high fine, no uncertainty.

In line with H5, the violation frequency is significantly lower for $F = 3600$ than for $F = 1200$

¹⁸Figure 3 illustrates the 131 out of 144 consistent decision profiles and the tests of Hypotheses 5 and 6 refer to these decision profiles. Test results are qualitatively the same when we consider all 144 decision profiles.

if there is legal uncertainty (i.e. in Figure 3, the solid black line is below the dashed black line; Wilcoxon signed ranks test, $p < 0.001$).¹⁹ As predicted by H6, for both fine sizes, the violation frequency is significantly lower in the absence of legal uncertainty than for high legal uncertainty (Wilcoxon signed ranks tests, $p < 0.001$ for both $F = 1200$ and $F = 3600$). In Figure 3, the dashed black line is well above the dashed gray line and the solid black line is well above the solid gray line.

Moreover, Figure 3 shows that players V often refrain from taking the money even when it would be payoff maximizing, thereby confirming findings by Gneezy (2005). For instance, in the presence of legal uncertainty and with a low fine, a monetary payoff maximizing player V would take the money if and only if the punishment frequency is below 80%, i.e. if at most 9.6 players J punish. However, in our experiment, only 22% of players V take the money if 9 players J punish. Similarly, in the absence of legal uncertainty and with a high fine, it is payoff maximizing to take the money whenever less than 4.8 players J punish, but only 9% of players V take the money if 4 players J punish. This could be driven by risk aversion or player V 's intrinsic aversion against norm violations as assumed in the theoretical model.

In sum, the evidence from our laboratory experiment is in line with all six hypotheses on partial effects of fines sizes and legal uncertainty on punishment and violation frequencies.

4.2 Comparison of partial and total effects of fine size and uncertainty

4.2.1 Punishment frequency

Our experimental design enables us to quantify total effects and to compare them to partial effects. We start with judges and then consider violators. Table 2 summarizes the insights from Proposition 2 on the punishment frequency.

We start with analyzing the total effect of the fine, for which our model yields a clear hypothesis:

H7. With legal uncertainty, the punishment frequency is lower for $F = 3600$ than for $F = 1200$.

By contrast, it depends on player J s' individual preferences to avoid type I relative to type II errors whether the total effect is larger or smaller than the partial effect. Table 3 and Table 4 summarize our data from the treatment in which players J conditioned their punishment decisions

¹⁹Also in line with our model, there is no significant impact of the fine size in the absence of legal uncertainty (Wilcoxon signed ranks test, $p = 0.191$).

Table 2: Theoretical predictions – total effects on the punishment frequency

	Total effect	Comparison to partial effect
Change in fine	$\frac{dp^*}{dF} < 0$	$\frac{dp^*}{dF} \begin{matrix} \geq \\ \leq \end{matrix} \frac{\partial p^*}{\partial F}$
Change in legal uncertainty	$\frac{dp^*}{dq} \begin{matrix} \geq \\ \leq \end{matrix} 0$	$\frac{dp^*}{dq} > \frac{\partial p^*}{\partial q}$

on the number of violating players V , while players V made unconditional, binary violation decisions. Since H7 refers to scenarios with legal uncertainty only, we restrict our attention to the two decision situations with $q = 50\%$. Taking the average over all three sessions, 9.3 out of 12 players V took the money for $F = 1200$, but only 5.3 for $F = 3600$ (see second row of Table 3).

The partial effect of the change in fine size on player J s' punishment behavior needs to be calculated under the counterfactual assumption that the violation frequency is independent of the fine size. Thus, we assume that the violation frequency is 9.3 not only for $F = 1200$, but also for $F = 3600$.²⁰ 75.8% of players J would have punished for $F = 3600$, compared to 79.4% for $F = 1200$. The difference is the partial effect of the fine size as shown in the fourth row of Table 3: If players J assumed that the higher fine had no impact on the violation frequency, then an increase in the fine size from $F = 1200$ to $F = 3600$ would lead to a rather small decline in the punishment frequency (3.6 percentage points).²¹

Table 3: Partial and total effects of the fine size on the punishment frequency

$q = 50\%$	$F = 1200$	$F = 3600$
Actual violation frequency	9.3 (78%)	5.3 (44%)
Punishment frequency given 9 violators	79.4%	75.8%
Partial effect	−3.6% <i>points</i>	
Punishment frequency with actual violation frequency	79.4%	48.5%
Total effect	−30.9% <i>points</i>	
Player J s' belief about violation frequency	6.8	4.6

The total effect is calculated by assuming that players J correctly anticipate the impact of the

²⁰Alternatively, we could assume that the number of violators is 5.3 not only for $F = 3600$ as observed in the data but also for $F = 1200$.

²¹We have already considered partial effects in Figure 2. Here, we calculate them explicitly as to compare them to total effects.

fine size on the violation frequency. Thus, we use the actual numbers of violations, 9.3 for $F = 1200$ and 5.3 for $F = 3600$, which, in line with H7, yields a rather large reduction in the punishment frequency of 30.9 percentage points.

While the precise numbers are specific to the parameters under consideration, the general message of our data is that the total effect is much more pronounced than the partial effect. Intuitively, players J who correctly anticipate player V s' behavior reduce their punishment frequency for two reasons: due to the higher fine size (direct effect) and due to the lower violation frequency (indirect effect). The last row in Table 3 shows that players J reasonably well anticipate the deterrent effect of higher fines, but underestimate the violation frequencies to some degree.

For the impact of legal uncertainty on the punishment frequency, our model yields no clear-cut prediction for the total effect itself, but for the comparison to the partial effect (see Table 2):

H8. The total effect of legal uncertainty on the punishment frequency is smaller than the partial effect.

Table 4: Partial and total effects of legal uncertainty on the punishment frequency

$F = 1200$	$q = 50\%$	$q = 0\%$
Actual violation frequency	9.3 (78%)	2.0 (17%)
Punishment frequency given 9 violators	79.4%	82.4%
Partial effect	3.0% <i>points</i>	
Punishment frequency with actual violation frequency	79.4%	70.6%
Total effect	-8.8% <i>points</i>	
Player J s' belief about violation frequency	6.8	3.4
$F = 3600$	$q = 50\%$	$q = 0\%$
Actual violation frequency	5.3 (44%)	1.0 (8%)
Punishment frequency given 5 violators	48.5%	70.6%
Partial effect	22.1% <i>points</i>	
Punishment frequency with actual violation frequency	48.5%	64.7%
Total effect	16.2% <i>points</i>	
Player J s' belief about violation frequency	4.6	1.7

Table 4 shows that, for both fine levels, legal uncertainty leads to a large increase in the violation frequency (for $F = 1200$ from 17% to 78% and for $F = 3600$ from 8% to 44% for an increase in legal uncertainty from 0 to 50%). To calculate the partial effects of an increase in legal uncertainty on

the punishment frequency, we again assume identical violation frequencies (78% for $F = 1200$; 44% for $F = 3600$) for both levels of legal uncertainty. For both fine sizes, legal uncertainty reduces the punishment frequency (partial effects of 3 or 22 percentage points, respectively). Interestingly, for $F = 1200$ the total effect of legal uncertainty has the opposite sign, i.e. legal uncertainty even leads to a higher punishment frequency (increase of 8.8%). Intuitively, if players J (correctly) anticipate that the violation frequency increases sharply in response to uncertainty, they have incentives to punish more often. The total effect goes in the opposite direction for $F = 3600$ (decrease of 16.2%). These seemingly contradictory total effects are actually well in line with our model which suggests that the sign of the total effect of legal uncertainty on the punishment frequency may be positive or negative – depending on whether the direct effect of legal uncertainty or the indirect effect (higher violation frequency) dominates player J s’ punishment decisions. Most importantly, in line with H8, the total effect is always smaller than the partial effect, documenting the existence of the countervailing indirect effect. Moreover, players J anticipate the overall violation frequencies surprisingly well.

Table 4 also offers information on how the total effect of the fine size on the punishment frequency differs in the absence and presence of legal uncertainty: with legal uncertainty, the punishment frequency for actual violation frequencies declines substantially from 79.4% given $F = 1200$ to 48.5% given $F = 3600$ (see rows 5 and 9 in the first column of Table 4). The corresponding reduction without legal uncertainty is smaller, from 70.6% to 64.7% (see last column of Table 4). Hence, even though the overall effect of the fine size on punishment under legal uncertainty might include some considerations that go beyond the aversion against error making discussed in our model, the small size of the effect for $q = 0$, which is less than a fifth of the effect under $q = 50\%$, suggests that these other considerations only play a subordinate role.

Overall, the experimental evidence supports the view that players J seek to anticipate adjustments in the behavior of players V so as to minimize their disutility from making type I and type II errors. In line with the Blackstone ratio, type II errors occur more frequently than type I errors: With legal uncertainty, 19.4% of players J make a type II error for $F = 1200$, and 27.8% of them make a type II error for $F = 3600$. The expected frequency of type I errors is 11.8% for $F = 1200$

and 13.2% for $F = 3600$.²²

4.2.2 Violation frequency

We now turn to comparing partial and total effects of fine size and legal uncertainty on the violation frequency. Based on Proposition 2, Table 5 summarizes our theoretical predictions.

Table 5: Theoretical predictions – total effects on the violation frequency

	Total effect	Comparison to partial effect
Change in fine	$\frac{d\phi^*}{dF} \begin{matrix} \geq \\ \leq \end{matrix} 0$	$\frac{d\phi^*}{dF} \begin{matrix} > \\ < \end{matrix} \frac{\partial\phi^*}{\partial F}$
Change in legal uncertainty	$\frac{d\phi^*}{dq} > 0$	$\frac{d\phi^*}{dq} \begin{matrix} \geq \\ < \end{matrix} \frac{\partial\phi^*}{\partial q}$

While our model does not provide an unambiguous prediction for the sign of the total effect for a change in the fine, it yields the following hypothesis:

H9. With legal uncertainty, the total effect of the fine size on the violation frequency is less pronounced than the partial effect.

Table 6 as well as Table 7 below present data from the treatment in which players V conditioned their violation decisions on the number of punishing players J , while players J made unconditional, binary punishment decisions for each combination of fine size and legal uncertainty. Since H9 applies to situations with legal uncertainty only, Table 6 focuses on the two decision situations with $q = 50\%$.

Table 6: Partial and total effects of the fine size on the violation frequency

$q = 50\%$	$F = 1200$	$F = 3600$
Actual punishment frequency	7.0 (58%)	3.7 (31%)
Violation frequency given 7 punishers	34.4%	12.1%
Partial effect	−22.3% <i>points</i>	
Violation frequency with actual punishment frequency	34.4%	27.3%
Total effect	−7.1% <i>points</i>	
Player V s' belief about punishment frequency	6.3	5.0

²²In our experiment, the incidence of type I errors is probabilistic. Given that player V has not violated but player J has decided to punish if the donation has disappeared, it depends on the random event that determines whether the donation disappears whether a type I error occurs.

To calculate the partial effect, we follow the same logic as for punishment frequencies. For $F = 1200$, 7 players J decide to punish, which induces a violation frequency of 34.4%. If 7 players J had also punished for $F = 3600$, the violation frequency would have been 12.1%. The difference of -22.3 percentage points thus captures the partial effect of an increase in the fine size. Using instead the actual number of punishing players J for $F = 3600$, 3.7, leads to a reduction of the violation frequency by only 7.1 percentage points, slightly less than one third of the partial effect. In line with H9, the deterrent effect of a higher fine is reduced as players V anticipate that the punishment frequency decreases in the fine (see last row of Table 6). Our strategy-method design reveals that differences of partial and total effects can be substantial: in our case, the increase in deterrence due to the higher fine when the judges' punishment behavior is anticipated is only one third of the partial deterrent effect. This result provides first empirical evidence in line with Andreoni's (1991) theoretical argument that higher fines are less effective in deterring violations as expected from the standard partial equilibrium perspective. Still, despite the ambiguous theoretical predictions on the sign of the total effect in Andreoni (1991) and in our model, our data suggest that the violation frequency decreases, even for a substantial increase in the fine and a high level of legal uncertainty.

For the total effects of legal uncertainty, our model predicts that, even when taking player J s' reactions to higher legal uncertainty into account, higher legal uncertainty reduces deterrence.

H10. The violation frequency is higher for $q = 50\%$ than for $q = 0\%$.

For a decrease in legal uncertainty from 50% to 0%, Table 7 shows total effects of -31.4 and -21.2 percentage points for the low and the high fine, respectively. Players V anticipate correctly (see information on beliefs in Table 7) that players J are less willing to punish in case of legal uncertainty and, in line with H10, uncertainty leads to more violations. This reinforces the view that legal uncertainty is worrisome not only from a justice point of view, but also from a deterrence perspective. Differences between partial and total effects, for which our theory yields no clear-cut prediction, are moderate.

5 Conclusion

We have analyzed the impact of fine size and uncertainty on the frequencies of punishments and legal infringements. In our theoretical model, we assume that judges have heterogeneous preferences

Table 7: Partial and total effects of legal uncertainty on the violation frequency

$F = 1200$	$q = 50\%$	$q = 0\%$
Actual punishment frequency	7.0 (58%)	10.3 (86%)
Viol. frequency given 7 punishers Partial effect	34.4% −28.3% <i>points</i>	6.1%
Viol. frequency with actual pun. frequency Total effect	34.4% −31.4% <i>points</i>	3.0%
Player Vs' belief about punishment frequency	6.3	10.0
$F = 3600$	$q = 50\%$	$q = 0\%$
Actual punishment frequency	3.7 (31%)	10.0 (83%)
Viol. frequency given 4 punishers Partial effect	27.3% −18.2% <i>points</i>	9.1%
Viol. frequency with actual pun. frequency Total effect	27.3% −21.2% <i>points</i>	6.1%
Player Vs' belief about punishment frequency	5.0	9.6

with regard to type I and type II errors, and that potential violators have different preference costs from violations. Based on these assumptions, we first confirm established theoretical results that higher uncertainty and higher fines reduce the punishment frequency when judges treat the violation frequency as exogenously given (*partial effects*). Analogously, higher uncertainty then increases, and higher fines reduce the violation frequency. In order to analyze *total effects* of fines and uncertainty, we extend the existing theoretical literature (see Andreoni, 1991 and Feess and Wohlschlegel, 2009) by fully taking the interdependency of judges' and potential violators' decisions into account. Moreover, we analyze comparative statics of both the violation and the conviction frequencies with respect to fine size and level of uncertainty. Results on total effects turn out to be less straightforward than the partial effects: while higher fines still reduce the punishment frequency and higher legal uncertainty still increases the violation frequency, the effect of the fine size on the violation frequency and the effect of legal uncertainty on the punishment frequency are theoretically ambiguous. However, both ambiguous effects are predicted to be less pronounced than the negative partial effects.

We then conduct a laboratory experiment to test our hypotheses on partial and total effects. The experimental results are fully in line with our theoretical predictions. Assuming that the

behavior of the other type of player is given (partial effects), higher fines reduce the violation and the punishment frequency, while legal uncertainty increases the number of violations and decreases the punishment frequency. Also in line with our theory, a higher fine leads to a significant reduction in the punishment frequency if and only if there is legal uncertainty. While partial effects have been studied in a variety of settings, the innovative part of our experiment is to determine total effects, and to compare them to partial effects. Our data on participants' beliefs about the behavior of the respective other type show that they do indeed anticipate changes in others' behavior in reaction to changes in fines or uncertainty. Moreover, the additional, indirect and sometimes countervailing effects of fine size and uncertainty that are predicted when assuming that violators and judges mutually anticipate their responses to fines and uncertainty do exist in our data. Comparing partial and total effects shows that differences can be sizable. Taken together, our results strongly underline the need to consider total effects - not only for economic models of legal decision making, but also in the design of real world institutions that aim at deterring illegal or unethical behavior.

In this latter regard, our findings suggest that higher fines are indeed likely to be less effective than expected from the standard partial equilibrium perspective since the punishment frequency is lower for higher fines. Moreover, the effect of legal uncertainty on the punishment frequency is less pronounced than the partial effect, because judges correctly anticipate that violation incentives increase in uncertainty. This is a countervailing effect that *ceteris paribus* reduces the risk of type I errors, and hence enhances the willingness to punish. Still, even when considering total effects, legal uncertainty substantially increases the violation frequency, which reinforces the view on the benefits of a well-functioning legal system.

Of course, one may (and should) wonder about the transferability of our results for real-life prosecution of unethical or criminal behavior. In real life, violations often have a strong moral connotation. In the experiment, we mimic a violation by the morally questionable act of taking money designated for donation to charity but consciously use a neutral wording otherwise, following standards in experimental economics. In our view, such an approach is perfectly adequate if it aims at testing the standard economic approach to violation and prosecution behavior as we do by testing Andreoni's (1991) and our theoretical predictions. The core of this economic approach relies on incentive effects triggered by fine sizes and uncertainty that are independent of all other

factors that may influence crime.²³ We are convinced that our findings have practical implications, for example for the frequent discussions on increasing fines to boost deterrence. Our results imply that, whenever there is legal uncertainty, the effectiveness of higher fines should at least be monitored closely since there might be a countervailing effect, an increased reluctance to actually impose the higher fines.

Acknowledgements

Financial support from the German Research Foundation (DFG) through SFB-TR 15 and by the Portsmouth Business School's Research Project Funds is gratefully acknowledged and did not influence study design or analysis of results. We thank Anne Mertens for programming and excellent assistance in implementing the experiment.

Appendix

A Proof of Proposition 2

If S_α and S_m are intervals, $\tilde{\alpha} \in S_\alpha$ and $\tilde{m} \in S_m$, then the Bayesian Nash equilibrium p^* and ϕ^* is given by the solution to the system of equations

$$p^* = p(\phi^*; F, q) = G(\tilde{\alpha}(\phi^*; F, q)) = G\left(\frac{\phi^*}{q(1 - \phi^*)} \frac{A}{F}\right) \quad (6)$$

$$\phi^* = \phi(p^*; F, q) = H(\tilde{m}(p^*; F, q)) = H(A(1 - p^*) - Fp^*(1 - q)). \quad (7)$$

Derivatives of Best-Reply Functions. For given ϕ , the impacts of F and q on p are

$$\frac{\partial p}{\partial F} = g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial F} \quad (8)$$

$$\frac{\partial p}{\partial q} = g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial q}. \quad (9)$$

²³Relatedly, Schildberg-Hörisch and Strassmair (2012) show that experimental results on deterrence remain qualitatively the same when using a morally loaded wording, in which decisions to take money are labeled as “stealing” and fines are referred to as “penalty” instead of “deduction points”. Abbink, Irlenbusch, and Renner (2002) also find no significant differences in bribing between framed and neutral instructions.

Similarly, for given p , the impacts of F and q on ϕ are

$$\frac{\partial \phi}{\partial F} = h(\tilde{m}) \frac{\partial \tilde{m}}{\partial F} \quad (10)$$

$$\frac{\partial \phi}{\partial q} = h(\tilde{m}) \frac{\partial \tilde{m}}{\partial q}. \quad (11)$$

Comparative Statics of Equilibrium Punishment and Violation Probabilities. The system of total differentials of the system (6) and (7) is

$$\begin{pmatrix} 1 & -g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial \phi} & -g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial F} & -g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial q} \\ -h(\tilde{m}) \frac{\partial \tilde{m}}{\partial p} & 1 & -h(\tilde{m}) \frac{\partial \tilde{m}}{\partial F} & -h(\tilde{m}) \frac{\partial \tilde{m}}{\partial q} \end{pmatrix} \begin{pmatrix} dp \\ d\phi \\ dF \\ dq \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (12)$$

Hence, the comparative statics are:

$$\frac{d\phi^*}{dF} = \frac{h(\tilde{m}) \left(\frac{\partial \tilde{m}}{\partial F} + \frac{\partial \tilde{m}}{\partial p} g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial F} \right)}{1 - h(\tilde{m}) \frac{\partial \tilde{m}}{\partial p} g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial \phi}} \quad (13)$$

$$\frac{dp^*}{dq} = \frac{g(\tilde{\alpha}) \left(\frac{\partial \tilde{\alpha}}{\partial q} + \frac{\partial \tilde{\alpha}}{\partial \phi} h(\tilde{m}) \frac{\partial \tilde{m}}{\partial q} \right)}{1 - g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial \phi} h(\tilde{m}) \frac{\partial \tilde{m}}{\partial p}} \quad (14)$$

$$\frac{d\phi^*}{dq} = \frac{h(\tilde{m}) \left(\frac{\partial \tilde{m}}{\partial q} + \frac{\partial \tilde{m}}{\partial p} g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial q} \right)}{1 - h(\tilde{m}) \frac{\partial \tilde{m}}{\partial p} g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial \phi}} \quad (15)$$

$$\frac{dp^*}{dF} = \frac{g(\tilde{\alpha}) \left(\frac{\partial \tilde{\alpha}}{\partial F} + \frac{\partial \tilde{\alpha}}{\partial \phi} h(\tilde{m}) \frac{\partial \tilde{m}}{\partial F} \right)}{1 - g(\tilde{\alpha}) \frac{\partial \tilde{\alpha}}{\partial \phi} h(\tilde{m}) \frac{\partial \tilde{m}}{\partial p}}, \quad (16)$$

Recall by Proposition 1 that $\frac{\partial \tilde{m}}{\partial p} < 0$, $\frac{\partial \tilde{m}}{\partial F} < 0$, $\frac{\partial \tilde{m}}{\partial q} > 0$, $\frac{\partial \tilde{\alpha}}{\partial \phi} > 0$, $\frac{\partial \tilde{\alpha}}{\partial F} < 0$, and $\frac{\partial \tilde{\alpha}}{\partial q} < 0$. Hence, the denominators of all right-hand sides are positive and larger than 1, and the numerator of the right-hand side of (15) is positive, and that of (16) is negative, which proves part (i) of the proposition. Furthermore, by (8), (9), (10) and (11), part (ii) of the proposition follow immediately. ■

References

ABBINK, K., B. IRLBUSCH, AND E. RENNER (2002): “An experimental bribery game,” Journal of Law, Economics, and Organization, 18(2), 428–454.

- ACEMOGLU, D., AND S. JOHNSON (2005): “Unbundling Institutions,” Journal of Political Economy, 113(5), 949–995.
- ANDERSON, C. M., AND L. PUTTERMAN (2006): “Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism,” Games and Economic Behavior, 54(1), 1–24.
- ANDERSON, L. R., G. DEANGELO, W. EMONS, B. FREEBORN, AND H. LANG (2017): “Penalty structures and deterrence in a two-stage model: Experimental Evidence,” Economic Inquiry, 55(4), 1833–1867.
- ANDREONI, J. (1991): “Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime?,” RAND Journal of Economics, 22(3), 385–395.
- ASSOCIATION OF CERTIFIED FRAUD EXAMINERS (2014): “Report to the Nations on Occupational Fraud and Abuse,” Discussion paper, http://www.acfe.com/uploadedFiles/ACFE_Website/Content/rtnn/2014-report-to-nations.pdf.
- BAUMANN, F., AND T. FRIEHE (forthcoming): “Crime and punishment under evidentiary uncertainty: Laboratory evidence,” Supreme Court Economic Review.
- BECKER, G. S. (1968): “Crime and Punishment: An Economic Approach,” Journal of Political Economy, 76(2), 169–217.
- CARPENTER, J. P. (2007): “The demand for punishment,” Journal of Economic Behavior & Organization, 62(4), 522–542.
- CASAL, S., F. FALLUCCHI, AND S. QUERCIA (2016): “Morals and markets: The case of ultimatum bargaining,” Discussion paper, University of East Anglia.
- CHALFIN, A., AND J. MCCRARY (2017): “Criminal deterrence: A review of the literature,” Journal of Economic Literature, 55(1), 5–48.
- COFFMAN, L. C. (2011): “Intermediation reduces punishment (and reward),” American Economic Journal: Microeconomics, 3(4), 77–106.

- COHEN, M. (2005): The costs of crime and justice. Routledge, New York, NY.
- DEANGELO, G., AND G. CHARNESS (2012): “Deterrence, expected cost, uncertainty and voting: Experimental evidence,” Journal of Risk and Uncertainty, 44(1), 73–100.
- DEANGELO, G., AND B. HANSEN (2014): “Life and death in the fast lane: Police enforcement and traffic fatalities,” American Economic Journal: Economic Policy, 6(2), 231–57.
- DURLAUF, S. N., AND D. S. NAGIN (2011): “Imprisonment and crime: Can both be reduced?,” Criminology & Public Policy, 10(1), 13–54.
- FEESS, E., AND R. SAREL (2018): Large Judicial Effort and the Appeal System: Theory and Experiment. Mimeo, Frankfurt School of Finance & Management.
- FEESS, E., AND A. WOHLSCHLEGEL (2009): “Why higher punishment may reduce deterrence,” Economics Letters, 104(2), 69–71.
- FEHR, E., AND U. FISCHBACHER (2004): “Third-party punishment and social norms,” Evolution and Human Behavior, 25(2), 63–87.
- FEHR, E., AND S. GÄCHTER (2000): “Cooperation and Punishment in Public Goods Experiments,” American Economic Review, 90(4), 980–994.
- FISCHBACHER, U. (2007): “z-Tree: Zurich toolbox for ready-made economic experiments,” Experimental Economics, 10(2), 171–178.
- FRANZONI, L. A. (1999): “Negotiated enforcement and credible deterrence,” Economic Journal, 109(458), 509–535.
- FRIESEN, L. (2012): “Certainty of punishment versus severity of punishment: An experimental investigation,” Southern Economic Journal, 79(2), 399–421.
- GNEEZY, U. (2005): “Deception: The role of consequences,” American Economic Review, 95(1), 384–394.
- GREINER, B. (2015): “Subject pool recruitment procedures: organizing experiments with ORSEE,” Journal of the Economic Science Association, 1(1), 114–125.

- HANSEN, B. (2015): “Punishment and Deterrence: Evidence from Drunk Driving,” American Economic Review, 105(4), 1581–1617.
- HARBAUGH, W. T., N. MOCAN, AND M. S. VISSER (2013): “Theft and deterrence,” Journal of Labor Research, 34(4), 389–407.
- HAREL, A., AND U. SEGAL (1999): “Criminal law and behavioral law and economics: Observations on the neglected role of uncertainty in deterring crime,” American Law and Economics Review, 1(1), 276–312.
- HASELHUHN, M. P., D. G. POPE, M. E. SCHWEITZER, AND P. FISHMAN (2012): “The impact of personal experience on behavior: Evidence from video-rental fines,” Management Science, 58(1), 52–61.
- HOLT, C. A., AND S. K. LAURY (2002): “Risk aversion and incentive effects,” American Economic Review, 92(5), 1644–1655.
- HUCK, S., AND M. KOSFELD (2007): “The dynamics of neighbourhood watch and norm enforcement,” Economic Journal, 117(516), 270–286.
- KHADJAVI, M. (2015): “On the interaction of deterrence and emotions,” Journal of Law, Economics, and Organization, 31(2), 287–319.
- KIRCHLER, M., J. HUBER, M. STEFAN, AND M. SUTTER (2016): “Market design and moral behavior,” Management Science, 62, 2615–2625.
- LEIBBRANDT, A., AND R. LÓPEZ-PÉREZ (2011): “The dark side of altruistic third-party punishment,” Journal of Conflict Resolution, 55(5), 761–784.
- LEVITT, S. D. (1998): “Juvenile Crime and Punishment,” Journal of Political Economy, 106(6), 1156–1185.
- NAGIN, D. S., AND G. POGARSKY (2003): “An experimental investigation of deterrence: Cheating, self-serving bias, and impulsivity,” Criminology, 41(1), 167.

- PNG, I. P. (1986): “Optimal subsidies and damages in the presence of judicial error,” International Review of Law and Economics, 6(1), 101–105.
- POLINSKY, A. M., AND S. SHAVELL (2000): “The Economic Theory of Public Enforcement of Law,” Journal of Economic Literature, 38(1), 45–76.
- RIZZOLLI, M., AND L. STANCA (2012): “Judicial errors and crime deterrence: theory and experimental evidence,” Journal of Law and Economics, 55(2), 311–338.
- SALMON, T. C., AND A. SHNIDERMAN (2017): Ambiguity in Criminal Punishment. Mimeo, Southern Methodist University.
- SCHILDBERG-HÖRISCH, H., AND C. STRASSMAIR (2012): “An experimental test of the deterrence hypothesis,” Journal of Law, Economics, and Organization, 28(3), 447–459.
- SCHLAG, K. H., AND J. J. VAN DER WEELE (2013): “Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality,” Theoretical Economics Letters, 3(1), 38–42.
- SCHRAM, A., AND G. CHARNESS (2015): “Inducing social norms in laboratory allocation choices,” Management Science, 61(7), 1531–1546.
- SHAVELL, S. (1985): “Uncertainty over Causation and the Determination of Civil Liability,” The Journal of Law & Economics, 28(3), 587–609.
- VOLOKH, A. (1997): “N guilty men,” University of Pennsylvania Law Review, 146(1), 173–216.